JIFs, giraffes, and a diffusion of culpability: A response to Osterloh and Frey's discussion paper on 'Borrowed plumes'

Article (Published Version)

Research Policy 49 (2020) 103868

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Discussion

JIFs, giraffes, and a diffusion of culpability: A response to Osterloh and Frey's discussion paper on 'Borrowed plumes'

Ohid Yaqub

SPRU, Science Policy Research Unit, University of Sussex, United Kingdom

1. A vestigial feature of the research system

The giraffe's laryngeal nerve is an absurdity. It connects the brain to the voice box, taking a circuitous route all the way down the neck and all the way back up again. Presumably, somewhere in the course of history, it got caught up with the development of other organs, and now takes a massive detour round the back of the aorta near the heart. It is anything but elegant, and was clearly "not meant to be, but just happens to be" (Gould, 1993, p76). Disentangling the nerve would involve the rearrangement of multiple organs simultaneously, and would be – to use policy parlance – a multi-stakeholder issue.

The journal impact factor (JIF) is an absurdity. It connects two weak bodies of understanding – citation theory and peer review theory – via a circuitous route through the research system. The JIF was supposed to assist librarians in curating their collection, and the underlying Citation Index[1] itself was supposed to help with searches and retrieval, but somewhere in the course of history, they both got caught up in an information flood. They are now used to rank selectively, a purpose for which they were certainly not designed. JIFs have ended up deeply entangled in modern research systems, and implicated in important pathologies of those systems.

Of course, the nerve remains functional for most giraffes, which is more than can be said about JIFs in the research system, so the metaphor is a strained one. Even a cursory look at the JIF reveals many reasons why it does not serve us well for ranking journals. It is a metric that is unforgiving of disciplinary differences and journal styles, and easily gamed by strategic authors, editors and publishers, to name just some of the issues (Archambault and Larivière, 2009; Braun, 2012; Martin, 2016). Perhaps most troubling is the idea that journal rankings may now be shaping the direction and content of science (Rafols et al., 2012; Muller and Rijke, 2017). Authors might complain about what is needed to get published in high-JIF journals, whilst editors might lament that submissions are becoming more homogenous. Journals certainly seem to shape and influence researchers' behaviour. Economists might not be willing to give their right arm for a publication in *American Economic Review* with its lofty stature, but the strength of their preferences imply they would at least be willing to sacrifice more than half a thumb (Attema et al., 2014)![2]

Yet, the metaphor does serve to illustrate a choice for addressing the current malaise in research policy. Should we 'muddle through', à la Lindblom (1959), tinkering with modified JIFs, using them in combination with other metrics, playing the arms race with those who seek to game them?[3] Or should we seek radical disentanglement from JIF, or indeed from journal rankings by any measure? Within this context, Osterloh and Frey (2020) do us an important service by asking why, for all its problems, journal rankings based on JIFs are still so influential. Moreover, they offer some suggestions for reform, some more radical than others.

In this note, I wish to focus on two points of difference between Osterloh and Frey, and myself. The first is the degree to which JIF lock-in can be explained exclusively by 'borrowed plumes'. The second is the degree to which their suggested reforms are sufficiently radical but too narrow in scope. Considering these two points, which turn out to be somewhat intertwined, will help to characterise what is really at stake. What is at stake is not so much how journal rankings affect academics and their careers ("top publications are decisive for academic careers"), but rather how well the research system interacts with and serves societal goals. Osterloh and Frey may well have identified one source of JIF lock-in, but I shall submit that there are other culprits, too, both within and beyond academia. Some are more culpable than others,

Part of the virtual Special Section on the Discussion Paper by Osterloh and Frey (2020).

E-mail address: o.yaqub@sussex.ac.uk.

[1] Originally, the Science Citation Index, subsequently joined by the Social Science Citation Index, and later the Arts and Humanities Citation Index.

[2] Are journals evolving in response to the authors' submission patterns? There is a hypothesis worth exploring: 'Top journals' may be growing their editorial boards and broadening their remit, so that they can cater for a more diverse range of interests under a single journal name, in what was previously accommodated through the auspices of multiple journals. Think of it, if you will, as a series of unintentional mergers and acquisitions, with the big journals hoovering up what they think will be the high-impact papers, even if they stretch the journal's topic remit. Journal communities might be less cohesive on the surface, but there may be more within-journal sub-communities; and a research community's publications start to become spread out across more journals as they are filtered by a priori perceptions of quality.

[3] For example, in response to carefully documented self-citation practices, intuitive policy recommendations can emerge, namely to develop modified journal-metrics and individual-metrics that exclude self-citations (Seeber et al., 2019; Wilhite et al., 2019).

https://doi.org/10.1016/j.respol.2019.103868
Received 2 July 2019; Received in revised form 31 July 2019; Accepted 9 September 2019
Available online 01 November 2019

readers may decide.

2. The degree to which JIF lock-in can be explained by 'borrowed plumes'

Osterloh and Frey (2020) begin by drawing our attention to one particular property of the JIF: skewed distributions. In general, a few papers in a given journal are much more highly cited than the rest of its papers. The authors of the rest of the papers, it is claimed, benefit from 'borrowed plumes' and are therefore likely to be reluctant to change. Blaming our inertia largely on free-riding authors, however, neglects a myriad of other kinds of 'borrowed plumes' that might exist and understates the broader systemic nature of JIF lock-in.

As a property of science, skewed distributions have been observed widely. Citation distributions across journals are skewed, citation distributions across papers within journals are skewed, citation distributions across authors are skewed, and citations distributions within authors' own portfolios of papers are skewed. We have known for a long time that authorship itself is also skewed; most authors publish relatively little while a few publish prodigiously (Lotka, 1926).

Arguably, much of the scientific enterprise might be said to run on 'borrowed plumes'. One might free-ride on supervisor pedigree, institutional brand-name, the reputation of co-authors, funding source prestige, or the generosity of peer reviewers, or indeed bask in the reflected glory of the field as a whole (witness the snake-oil salesmen rushing to call themselves economists amongst politicians). The rewards are so extremely unevenly distributed that one author has even drawn a parallel here between the academic community and drug gangs (Afonso, 2014).[4]

Conversely, in the Mathew Effect, the plumes can be blown in the reverse direction, such that the superstars reap cumulative advantage from their earlier work, and from the toils of technicians, graduate students, and administrative assistants. They probably garner more than their 'fair share' of citations too, once they have acquired the aura of elite high-status (Azoulay et al., 2013). Hence, to take umbrage at this one particular set of borrowed plumes arising from the JIF, I would suggest, distracts attention from the fact that there is already in place an extremely stratified system of rewards and incentives for researchers - journal rankings merely uphold and exaggerate these inequalities. Moreover, as a verdict on the acquiescence of authors of the papers who get cited less than their respective JIF, it seems a rather harsh one, and sets a stern threshold for who should be held complicit.

Where to start the list then? Perhaps a great source of responsibility lies with those who incorporate the JIF into their broader rankings of journals, researchers, departments and entire universities. Rankings can introduce dysfunctionalities over time as measures become targets[5] and, irrespective of whether they are initially misrepresentative or not, they can gain credence over time, as Mertonian self-fulfilling prophecies kick in:

> "Law school faculties and the smart administrators all go, 'This is a bunch of hooey. We don't care about this.' Until they drop and the board of trustees says, 'Hey, you're dropping; why should we give you more money?' and the board of visitors say, 'Man, your school's really going to pot. You haven't changed a thing… big changes need to be made here.' And your monetary support from alumni say, 'Well, I'm not sure I want to support a school that's going in the wrong direction,' and your money starts to dry up. And you think, 'we can't afford to lose funding, or else it *will* spiral downhill and we will be a worse law school. So keeping up numbers is key." (Quoted in Espeland and Sauder, 2007, p13)

Might the new breed of university managers be responsible? If rankings are thought to be credible, and capable of controlling the fate of a university, school or department, those rankings may be incorporated directly into the 'key performance indicators' of university managers, or even codified into their employment contracts and remuneration packages. The cadre of managers may then be tempted to do what it takes to maintain and rise up the rankings in order to execute their duties. Implicit pressures to encourage manuscript submissions to 'top journals', which researchers might already be feeling anyway, can become explicit bonus payments for publication in certain select outlets, or turn into the threat of dismissal without publication in sufficiently high-JIF journals.

Moreover, if there are activities in the university or school that have nothing to do with publishing in high-JIF journals, and are not included in the rankings, management might be tempted to withdraw support for them. Some might say, just create a new set of rankings that measure what is desired, for example, a renewed focus on impact and service. Putting aside the diversity and politics of what might constitute impact and service, such an argument would overlook the ways in which rankings encourage homogeneity. Schools become more like what is measured. Public policy schools, business schools, law schools, and medical schools start to the look the same as each other, their character differences ironed out into top-tier and lower-tier schools. What might have made a school distinctive, survives only if it can be framed as a source of competitive advantage.

Rankers might remind us that they, too, have a responsibility to their audience, that there is a market for their service, warts and all. If they did not offer it, someone else would probably step into the market. So the users and consumers of rankings are enveloped in this mess, too. Those who might defer to rankings – funders allocating grants to researchers; committees deciding on recruitment, promotion and prizes; students and their parents choosing where to study; governments limiting where their scholarships can be used – suddenly become drawn into a vast web of diffused evaluation culpability.

Even those who might refrain from using such rankings do not escape unsullied, since one only has to be suspected of taking a sneaky peek at them, for it to prompt reactive behaviour. Witness, for example, the UK Research Excellence Framework panels insisting that they do not allow journal name, let alone journal rankings, to affect their judgments. Given that the REF's funding allocation and status rewards are now skewed even more heavily towards quality of publications rather than quantity, anxiety about how quality will be judged is to be expected. The correlation between REF peer review and journal rankings may be weak for now (HEFCE, 2015; p32), but there is a danger of this ending up as a self-fulfilling prophecy.

More broadly, the pressure to select and concentrate has become stronger, yet the appetite to engage seriously in evaluation, and to fund it properly, has not grown with the same fervour. The shift from steadily growing science to steady-state science has resulted in competition for ever more limited resources. The decline in institutional funding has left universities vying for tuition fees and its researchers jostling for short-term project funding. Meanwhile, as the number of manuscripts being generated and submitted for publication has increased, no doubt in part because this is a performance indicator, evaluation and selection mechanisms are coming under greater stress. This is perhaps more acute than it needs to be because publishers under-fund evaluation and selection, preferring to rely on the freely donated time of editors and peer reviewers.

The appetite for investing in evaluation ought to be stronger than it is, not just among publishers but more generally, too. Under such

[4] However (and I say this without irony), this claim does not appear to have got past peer review in its original form in that this working paper has yet to be published.

[5] This is widely recognised. See for example, Goodhart's Law (quoted in Martin, 2016; p6); Campbell's Law (1979, and quoted in Espeland and Sauder, 2007, p3); metrics as 'distortions' of university management (Muller, 2018, p23 and p67); 'performativity' in markets (Callon, 1998; Mackenzie, 2006); and 'audit culture' in accounting and actuary (Porter, 1996, p89; Power, 1997, p15).

conditions, it becomes tempting to reach for evaluation 'short-cuts', or to use Dr Wooding's term, heuristics. As Wooding (2020) suggests, there is simply too much at stake for research to be left unevaluated, so there will always be a need to ensure people do not feel left at sea, largely unguided as to which heuristics to use.

3. The degree to which the suggested reforms are radical but narrow

Osterloh and Frey's suggestions for reform are radical but narrow: open post-publication review; publication of citation distributions; 'take it or leave it' publication decisions (i.e. no revisions); and randomisation (whenever there is reviewer disagreement). They are admirably weighed up with a series of carefully thought-through advantages and disadvantages, which do not need repeating here (not least because the arguments there have already been sharpened through peer review!). Suffice it to say that the first two suggestions are already in partial operation, and the third speeds up review but somewhat undervalues the contributions that reviewers can make in suggesting improvements (Starbuck, 2003; Casnici et al., 2017).

The fourth option of partial randomisation triggers a number of questions, some of which have been answered but others not: Would this encourage yet more speculative manuscript submissions, knowing that they need only satisfy a single reviewer to enter a lottery? Would this change the behaviour of reviewers if they knew they could unilaterally consign a manuscript to a lottery? Are editors no longer to play a role in at least trying to understand the sources of author-reviewer disagreements or reviewer–reviewer disagreements, and the different criteria by which reviewers may be evaluating the quality or originality of the contribution? Does this relegate editors to being mere administrators and processors of papers, rather than active custodians of their community's publishing activities?

I am not sure how to answer these questions, but they do need careful consideration before a major overhaul. Swapping the biases and challenges of editing a journal for randomising a significant portion of a journal does seem to imply a rather strong indictment of editorial judgement. It remains to be seen whether this is what readers want, or what the editors who serve them want. At present, there is little incentive for publishers and editors to even experiment with Osterloh and Frey's proposition. Oswald (2020) lays out one rationalisation, but it involves a rather literal take on citations (requiring one to value citations for citations' sake, "admittedly just one criterion"). The extent to which reviewers' and editors' view of high-impact is basically equated to potential for citations is unknown, but some journal rankings certainly apply reductive pressure in that direction. If Oswald's starting assumptions are not valid now, they may yet become valid, when chasing citations becomes the new orthodoxy.

If unorthodox views are in danger, however one chooses to define that, editors should in principle be able to over-rule reviewers and publish anyway. When they don't, unorthodox ideas in the past have sometimes been reason for setting up a new journal. Some journals can establish a leading reputation quite quickly; this journal is coming up to fifty years, but I'm sure many will be able think of younger examples that are also very good. Today's tendency to rank journals probably makes striking out with a new journal much harder than in years gone by, but despite this, it has not stopped the continued birth of new journals.

It remains unclear whether it is journal ranking by JIF or journal ranking by any measure that is objectionable to Osterloh and Frey. Compared to the DORA (2012) declaration and the Leiden Manifesto (Hicks et al., 2015), which are setting new social norms in science and its institutions (a set of standards that can be used to shame and revere in equal measures is perhaps nowhere more useful than on social media platforms); and compared to the Metric Tide report (Wilsdon et al., 2015), which is focussing policymakers' attention on the pernicious effects of the JIF and the wide ranging implications for research assessment more broadly, the Osterloh and Frey reforms risk appearing to be an esoteric debate within the ivory tower about an arcane system of promotions and publishing. Rather than taking aim at authors-cited-below-their-JIF-scores and their journal editors, the intended audiences of the DORA declaration, the Leiden manifesto and the Metric Tide have deliberately been much wider.

These reforms are broader in scope, but also less radical in that they do not seek to dismantle the journal system that has evolved, and do not try to break free from journal-ranking in its entirety. Rather, they implicitly accept a weak form of journal hierarchy (for example, at the very least seeking to distinguish scholarly journals from a growing range predatory journals whose only editorial concern is to extract rents from the academic market by charging gullible authors, or those desperate to get a 'publication' regardless of the journal). They don't ignore the arms race referred to earlier, between those who develop ("a basket of") metrics and those who seek to game them; they seek for it to be governed, via multiple stakeholders, with all the perils that may entail, in the hope that diversity might be respected amidst a weaker form of hierarchy. The JIF might yet be governed into an irrelevance. Reform need not be radical and risky; there is merit in having broad reach.

The latest in this stream of work has been the suggestion of a governing body of journals (perhaps to fill the vacuum left by those preoccupied with extraordinary rent-seeking).

> "All stakeholders in the system share responsibility for the appropriate construction and use of indicators, but in different ways. We therefore suggest the creation of an inclusive governing organization that would focus on journal indicators." (Wouters et al., 2019; p622)

This seems both feasible and promising. Perhaps one of the first items on their agenda ought to be supporting further work on citation theory and peer review theory, if only to help maintain that knowledge base and garner wider appreciation for the literature that already exists. For example, known variations in citation practices are too often overlooked (Elkana et al., 1978; Cozzens, 1989; Wouters, 1999). And, the expectation that peer review should yield consensus seems far too strong, given how differently we know that people can read the same document (c.f. the three responses here to the same Osterloh and Frey manuscript) (Darnton, 1986; Chubin and Hackett, 1990; Csiszar, 2016). We must endeavour to offer more than superficial answers to these questions: What are journals for? What is peer-review? And what is a citation?

Our answers will require continual updating if histories of reading, print culture, journals, peer review, citation culture are anything to go by. So we do need a forum for discussing what we think a citation is, what journals are for, and what peer review does. Such discussions could inform, and be informed by, problems facing research evaluation practitioners. For example, evaluators grapple with the unit of analysis (which evaluation techniques are appropriate for which level of research aggregation?) and the degree of commensurability (should a citation from a student be weighted the same as one from a Nobel laureate, or citations from a blog the same as one from a newspaper?),[6] in addition to the perennial concerns about the purpose of evaluation (summative, formative, accountability, and so forth) (Marjanovic et al., 2017).

This is, of course, not the first time that communities have encountered the sensation of information overload and the need for selection systems (Blair, 2010). The invention of the printing presses might well be considered as nothing less than an information revolution by some (Eisenstein, 1979), but it was not without a concomitant evolution in the governing institutions that shaped publishing, reading and censorship (Johns, 1998). The idea of a governing body is

[6] Citations can be 'normalised' to account for variations, like field of study or career stage, but the underlying question is theoretical: what is 'normal'?

reminiscent (to me at least, since I have not spoken in respect of this to any of the authors of Wouters et al., 2019) of the Company of Stationers in its heyday, whose royal charter in 1557 set out a definition of "proper conduct":

> "To be a Stationer meant adopting a distinctive cultural identity… that conditioned how every individual Stationer perceived his or her knowledge, conduct and action. By comparison, categorization into 'bookseller', 'printer' or 'wholesaler' – let alone 'publisher' or 'editor' – was relatively unfamiliar. Such terms possessed nothing like the resonance of 'Stationer'. [They] acted first and foremost as Stationers and it was as Stationers that they and their actions were judged… The decisions structuring print culture were overwhelmingly Stationers' decisions, arrived at by reference to Stationers' perspectives. Their interests and practices therefore had direct implications for virtually all learned activities." (Johns, 1998; p59).

To sum up, explanations for JIF lock-in ought to go beyond 'borrowed plumes'. JIFs, journal rankings, and other forms of ranking, spin a web that catches many actors. It seems sensible to focus on reform that has potential for wide reach. Randomisation in the face of JIF may carry unintended consequences and may not succeed in dislodging the desire for journal rankings by some other measure(s). It should wait until we have more widely appreciated theory on peer review and citation, more inclusive governing bodies that can wield some influence over rankings and their users, and a stronger appetite for investing in evaluation. For the time being at least, we might learn to live with JIFs by engaging with a wide range of stakeholders, and without the need for such a radical shake-up of the journal system. It is not just giraffes who have managed to live with the laryngeal nerve – humans have too.

Declaration of Competing Interest

None.

Acknowledgements

I acknowledge funding from European Research Council Starting Grant 759897. I thank Ben Martin, the Editor of this (virtual) Special Section, for helpful comments. All errors are my own.

References

Afonso, A., 2014. How academia resembles a drug gang. Retrieved July 30, 2019, Available at https://doi.org/10.2139/ssrn.2407748.
Archambault, E., Larivière, V., 2009. History of the journal impact factor: contingencies and consequences. Scientometrics 79 (3), 635–649.
Attema, A.E., Brouwer, W.B.F., Van Exel, J., 2014. Your right arm for a publication in AER? Economic Enquiry 52 (1), 495–502.
Azoulay, P., Stuart, T., Wang, Y., 2013. Matthew: effect or fable? Manage Sci 60 (1), 92–109.
Blair, A., 2010. Too Much to Know: Managing Scholarly Information before the Modern Age. Yale University Press, New Haven.
Braun, T., 2012. Editorial. Scientometrics 92 (2), 207–208.
Callon, M., 1998. The Laws of the Markets. Blackwell, Oxford.
Casnici, N., Grimaldo, F., Gilbert, N., Dondio, P., Squazzoni, F., 2017. Assessing peer review by gauging the fate of rejected manuscripts. Scientometrics 113 (1), 533–546.
Chubin, D., Hackett, E., 1990. Peerless Science: Peer Review and U. S. Science Policy. SUNY, New York.
Cozzens, S., 1989. What do citations count? Scientometrics 15 (5–6), 437–447.
Csiszar, A., 2016. Peer review: troubled from the start. Nature 532, 306–308.
Darnton, R., 1986. First steps toward a history of reading. Aust. J. Fr. Stud. 23 (1), 5–30.
DORA(San francisco declaration on research assessment), 2012. Retrieved July 30, 2019, from https://sfdora.org.
Eisenstein, E., 1979. The Printing Press as an Agent of Change. CUP, Cambridge.
Elkana, Y., Lederberg, J., Merton, R.K., et al., 1978. Toward a Metric of Science: The Advent of Science Indicators. Wiley, New York.
Espeland, W., Sauder, M., 2007. Rankings and reactivity. Am. J. Soc. 113 (1), 1–40.
Gould, S., 1993. Eight Little Piggies. Norton, New York.
HEFCE, 2015. The Metric Tide: Correlation analysis of REF2014 Scores and Metrics (Supplementary Report II to the Independent Review of the Role of Metrics in Research Assessment and Management). HEFCEhttps://doi.org/10.13140/RG.2.1.3362.4162.
Hicks, D., Wouters, P., Waltman, L., et al., 2015. Bibliometrics: the leiden manifesto for research metrics. Nature 520, 429–431.
Johns, A., 1998. The Nature of the Book. CUP, Chicago.
Lindblom, C.E., 1959. The science of "muddling through". Publ. Adm. Rev. 19 (2), 79–88.
Lotka, A., 1926. The frequency distribution of scientific productivity. J. Washington Acad. Sci. 16 (12), 317–324.
MacKenzie, D., 2006. An engine. Not a Camera: How financial Models Shape Markets. MIT Press, Cambridge, MA.
Marjanovic, S., Cochrane, G., Robin, E., et al., 2017. Evaluating a complex research capacity-building intervention: reflections on an evaluation of the African Institutions Initiative. Evaluation 23 (1), 80–101.
Martin, B., 2016. Editors' JIF-boosting stratagems –which are appropriate and which not? Res. Policy 45 (1), 1–7.
Muller, J.Z., 2018. The Tyranny of Metrics. Princeton University Press, NJ.
Muller, R., Rijke, S., 2017. Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences. Res. Eval. 26 (3), 157–168.
Osterloh, M., Frey, B., 2020. How to avoid borrowed plumes in academia. Res. Policy 49, 103831 (this virtual Special Section).
Porter, T., 1996. Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press, NJ.
Power, M., 1997. The Audit Society: Rituals of Verification. Oxford University Press, Oxford, UK.
Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., Stirling, A., 2012. How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. Res. Policy 41 (7), 1262–1282.
Seeber, M., Cattaneo, M., Meoli, M., et al., 2019. Self-citations as strategic response to the use of metrics for career decisions. Res. Policy 48 (2), 478–491.
Starbuck, W., 2003. Turning lemons into lemonade: where is the value in peer reviews? J. Manag. Inquiry 12 (4), 344–351.
Wilhite, A., Fong, E., Wilhite, S., 2019. The influence of editorial decisions and the academic network on self-citations and journal impact factors. Res. Policy 48 (6), 1513–1522.
Wilsdon, J., Allen, L., Belfiore, E. et al. (2015). The metric tide: report of the independent review of the role of metrics in research assessment and management. doi:10.13140/RG.2.1.4929.1363.
Wouters, P., 1999. Citation Culture. UAP, Amsterdam.
Wouters, P., Sugimoto, C., Larivière, V., et al., 2019. Rethinking impact factors. Nature 569, 621–623.